DOCUMENT RESUME

ED 430 183 CG 029 270

AUTHOR Hummel, Thomas J.; Lichtenberg, James W.
TITLE Interpretive Error in Psychological Testing.
PUB DATE 1999-04-20
NOTE 22p.; Paper presented at the Annual Conference of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Ambiguity; Counseling; Evaluation; *Hypothesis Testing; *Language; *Measures (Individuals); Models; *Psychological Testing; Psychometrics; *Test Interpretation

ABSTRACT

When evaluating the utility of a psychological test for clinical decision making, both the psychometric properties of the test (i.e., the reliability and validity of the instrument) and the ambiguity of the language by which test results are interpreted or communicated need to be considered. Although each has been studied independently, to date the two have not been related. This paper discusses each of these sources of "interpretive error" with the goal being the development of a model that could systematically relate these two sources of error to the process and outcome of test interpretation. In an example using this new model together with optimistic good-hearted assumptions favorable to tests, it was found that the effect of ambiguity was to more than double the probability that those who test positive will be incorrectly classified. It is suggested that the practice of using non-numeric statements of likelihood be stopped. Instead, results of tests should be presented in tabular form with one row for each diagnostic category. Examples of these tables are provided. It is further noted that there is a significant professional need to move toward empirically supported expertise in test interpretation. Contains 24 references. (Author/MKA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 430 183

Interpretive Error in Psychological Testing

Thomas J. Hummel
University of Minnesota

James W. Lichtenberg
University of Kansas

Paper presented as part of a symposium on "Research and Clinical Issues in Test Interpretation" (S. Jenkins, Chair) at the annual meeting of the American Educational Research Association, April 20, 1999, Montreal, CANADA.

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. LICHTENBERG
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

CG029270

Abstract

When evaluating the utility of a psychological test for clinical decision making, both the psychometric properties of the test (i.e., the reliability and validity of the instrument) and the ambiguity of the language by which test results are interpreted or communicated need to be considered. Although each has been studied independently, to date the two have not been related. In this paper we discuss each of these sources of "interpretive error" with the goal being the development of a model that could systematically relate these two sources of error to the process and outcome of test interpretation. In an example using our model together with optimistic good-hearted assumptions favorable to tests, we found that the effect of ambiguity was to more than double the probability that those who test positive will be incorrectly classified.

Interpretive Error in Psychological Testing

The move toward empirically supported interventions in counseling psychology is a move to push counselors to justify with data their treatment of clients (Elliott, 1998; Kendall, 1998). If counselors' treatments are differentially effective depending on the client's problem, as published criteria for empirically supported interventions would imply (Chambless & Hollon, 1998), then diagnosis and the differentiation of client problems is required. Clearly, having empirically supported interventions without empirically supported diagnoses would go no where, for although a counselor would have a set of treatments for different client problems, the counselor would have no way to determine which clients had which problems.

Within the area of psychological assessment, it is generally accepted that the correlations between test scores and behavior is less than perfect. A cursory review of recent texts in this area suggests that these correlations generally range between .3 and .7 for academic measures (depending on variables entered into the prediction equation) and average between.3 and .5 for personality measures. With respect to personality measures, this translates to over 75% error variance. The error attributable to low test (predictor) to behavior (criterion) correlations—which frequently is indexed using the standard error of estimate—together with low clinical base rates, contributes to "prediction error" in psychological testing; and the magnitude of such error has led some (e.g., Hummel, 1999) to question the usefulness of tests for in making clinical decisions.

Another independent source of error is in how the results of testing are communicated to and understood by clients. Previous research (see Lichtenberg & Hummel, 1998 for a review) has shown there to be significant variability in the way that individuals interpret the probabilistic

statements that are used to describe and communicate test results. Despite a common formal training in psychometrics and in the use of specific tests, the evidence would suggest that counselors as recipients of, or as communicators of, test interpretations do not share common (quantitative) meanings for the probabilistic expressions that are used in providing test interpretations.[1]

When evaluating the utility of a psychological test for clinical decision making, both the psychometric properties of the test (i.e., the reliability and validity of the instrument) and the ambiguity of the probabilistic language by which test results are interpreted and communicated need to be considered. Although each has been studied independently, to date the two have not been related. In this paper we discuss each of these sources of "interpretive error" with the goal being the development of a model that could systematically relate these two sources of error to the process and outcome of test interpretation.

The "Fuzzy Language" of Test Interpretation

Counselors, like many other professionals share the occupational requirement of having to deal with uncertainty and communicating it to others. The counselor predicts that the client is *likely* to find a particular occupation or course of study satisfying, or that the child is ***unlikely*** to succeed in a regular classroom setting, or that the client ***may*** attempt suicide, or that the client *probably* was abused as a child, or that the client ***almost certainly*** is a child molester, or that a particular intervention is *likely* to be effective, or that it is ***possible*** that the client will become violent, or s/he states that the client *occasionally* has flashbacks, or that individuals with this profile are ***fairly common***.

The application of psychological testing is in large part an attempt to derive probabilistic statements regarding the likelihood of occurrence of client states, choice outcomes, situational

antecedents, and behavioral outcomes. Grounded in psychometric theory, psychological tests are an attempt to quantify these probabilities, and directly or indirectly, psychological test interpretation—whether done clinically or mechanically (Goldman, 1973)—is an attempt to translate and express those probabilities into words rather than numbers.

Test interpretations, written or oral, may be provided to counselors by other professionals or test scoring services, or they may be made to clients, sanctioners of services (e.g., parents, the courts, employers), fellow professionals, or others with a legitimate need and right to know. How the recipients of such interpretations translate these qualitative descriptions of behavioral probabilities into numerical estimates of attributes or outcomes is unclear; although there is considerable evidence drawn from literature outside of counseling psychology to suggest that it would be unwise to assume that they do so uniformly (Budescu & Wallsten, 1985; Reagan, Mosteller & Youtz, 1989; Sutherland, et al., 1991). Despite a common formal training in psychometrics and in the use of specific tests, prior research on the subjective and communicative meaning of probabilistic phrases (e.g., Bass, Cascio, & O'Connor, 1974; Beyth-Marom, 1982; Brun & Teigen, 1988; Budescu & Wallsten, 1985; Clarke, Ruffin, Hill, & Beamen, 1992; Foley, 1959; Johnson, 1973; Lichtenstein & Newman, 1967; Ness, 1995; Simpson, 1944, 1963; Wallsten, Budescu, Rapoport, Zwick & Forsyth, 1986) would suggest that counselors as recipients or communicators of test interpretations are unlikely to share common (quantitative) meanings for the probabilistic expressions they use in test interpretations. Nevertheless, the extent of the variance in counselors' understanding of probabilistic language used within the context of psychological test interpretation previously has not received much consideration.

In a previous study, we (Lichtenberg & Hummel, 1998) focused on how counselors receiving standard, computer-generated MMPI and MMPI-2 interpretive reports about a client might variously understand the probabilistic language used in the reports. Within these narrative interpretive reports, the use of terms such as "probable" or "possible" with respect behavioral or characterological correlates of test scores is intended to convey a meaning or interpretation that implies a certain degree of probability. In a reciprocal fashion, however, the counselor as receiver of the expression interprets or understands a certain degree of probability associated with the words used in the report. Confusion, or at least miscommunication, is likely to result if the meaning attached to a probability expression in a report is significantly different from the meaning assumed by the recipient of the expression. If, for example, "probable" means "about 50% of the time" to the person or service preparing the report and "about 80% of the time" to the recipient of the report, the differences in the subsequent clinical decisions and courses of action implied by these different meanings may be significant.

To the extent that counselors' understanding of the probabilistic language in test reports differs from that intended by the preparer of the report, there is interpretive error compounding already existing measurement and prediction error. Such error is likely to be compounded still further as counselors, using probabilistic language, offer interpretations based on their understanding of the probabilistic language in reports they receive to their clients.

The implications of such language imprecision can be significant for all concerned, as such information is used to make important diagnostic and subsequent treatment decisions concerning the person tested. An intervention or course of action may be taken with respect to a client based on diagnostic determinations that in turn were based on the fuzzy language of narrative test interpretation. A client may be hospitalized (or released from hospitalization) on

the strength of the interpretation provided. A student may be advanced or held back in school based on probability estimates provided regarding the child's likelihood of success in the next grade. Parents may seek or terminate special education services based on their belief regarding the likely benefit of such services—a belief shaped by the interpretation of their child's psychological testing. A client may choose to pursue a particular college major based on the understanding that she would be "likely satisfied" with a particular career.

The results of our study (Lichtenberg & Hummel, 1998) suggested that counselors are far from uniform in their understanding of the probabilistic language used in narrative test interpretations. As evidenced by the degree of variability in the numerical probability ratings assigned to the probabilistic language contained in actual MMPI and MMPI-2 narrative reports and by our finding of significant overlap in the numerical meaning attributed to this language (despite significant difference in the average numerical probability ratings assigned these expressions), our results suggested that verbal probability expressions appear to be a poor tool to convey test interpretations.

Clinical Hypothesis Testing

When a client's diagnosis and subsequent treatment are determined using tests, then it is reasonable to inquire as to the decision model the counselor uses when making such decisions. In particular, it is reasonable to ask how diagnostic error is controlled. Hummel (1990, 1999) has proposed a measurement-based hypothesis testing procedure called *clinical hypothesis testing*. The model is based on concepts similar to those used in inferential statistics—particularly in statistical hypothesis testing. It allows counselors to test specific clinical/diagnostic hypotheses about clients with whom they work.

The model involves seven basic, but essential, steps:

1. State a null hypothesis that the client is not a member of the diagnostic group in question (e.g., the client will not become a delinquent or the client is not depressed).
2. Determine the criterion variable and rule for inclusion in the diagnostic category (e.g., what set of conditions must exist for the client to be considered "delinquent" or "depressed").
3. Consider the consequences of misclassification or misdiagnosis.
4. Set the maximum probability of incorrectly classifying the client as being a member of the diagnostic category (similar to setting an α-level in statistical hypothesis testing).
5. Compute a critical value on the predictor variable (similar to a critical value in statistical hypothesis testing; i.e., when a result is as or more extreme than the critical value, one rejects the null hypothesis).
6. Compare the client's score on the predictor variable to the critical value and make a decision (similar to testing a statistical hypothesis by comparing the obtained value of a test statistic to a critical value).
7. (Optional) Determine the long-run probabilities of true positives, false positives, true negatives and false negatives that would result if the critical value were used over time with all clients (somewhat similar to Type I and Type II errors in statistical hypothesis testing). (Hummel, 1999, p. 63)

When using this approach for classifying or diagnosing clients, an error is made when a client is said to be in a diagnostic group when in truth he or she is not (e.g., a client is classified as depressed when she or he is not—in testing terminology, a "false positive"). This is similar to a Type I error in statistical hypothesis testing—when a true null hypothesis is rejected. A clinical diagnostic error also occurs when a client is actually in a diagnostic group but is not

assigned to that group by the counselor's clinical decision rule (e.g., when a diagnostic decision is made that the client is not depressed when in fact he or she is—in testing terminology, a "false negative"). This is similar to a Type II error in statistical hypothesis testing—when a false null hypothesis is incorrectly retained.

Controlling "Type I" and "Type II" error in the diagnosis of client is a professionally responsible goal. And although both types of diagnostic errors are have important clinical implications, it is Hummel's position (Hummel, 1990, 1999) and ours in this paper that controlling the probability of false positives ("Type I error"), that is rendering a formal diagnosis when one is not appropriate, should be a primary goal.

The details of Hummel's Clinical Hypothesis Testing model (CHT) have been detailed in other papers and will not be reintroduced here. Suffice it to say that it is assumed that for any given diagnostic situation there is a maximum probability of error acceptable to the counselor. As noted previously, setting this probability is analogous to setting an α-level in statistical hypothesis testing. As with the setting of α-levels, the value chosen is a personal judgment (although in statistical hypothesis testing a value of 0.05 is conventional). As demonstrated in previous papers, however, given the validity coefficients of available tests and the base rates of typical diagnostic groups, a decision to control error at the 0.05 level would result in abandoning tests for diagnostic purposes. This is because the proportion of clients testing positive would be so small that the cost of testing could not be justified. For this reason, counselors may need to accept a more moderate likelihood of error and a higher risk of a false positive.

The Integration of "Fuzzy Language" and the CHT Model

The Clinical Hypothesis Testing model has as its goal the control diagnostic error—specifically the misclassification of clients into diagnostic categories. Its utility derives

from the application of classification "cut scores," the test user's acceptable level of decision-error, the base rates of the diagnostic classifications under consideration, and the criterion-related test validity. But as noted earlier in this paper, the information that counselors have to consider when making diagnostic decisions about clients is not always so neatly quantitative. Rather it frequently consists of a narrative report of the likelihood of various client thoughts, feelings and behaviors that have been found to be associated with certain diagnostic classifications. How the ambiguity of this "fuzzy language" may be considered within the CHT model is the goal of the remainder of this paper. Here we attempt to outline a model that could systematically relate these two sources of error to the process and outcome of test interpretation.

In using tests for clinical diagnosis, suppose the criterion, y, and the predictor, x, are jointly distributed according to a standard bivariate normal distribution.[2] Further suppose that the criterion and predictor are positively related and that higher scores on the criterion are more problematic, i.e., the clinical group is in the upper tail of the criterion distribution. In CHT, there are two characteristics of the population that are important, the base rate, ϕ, of the clinical group and the validity coefficient, ρ. There are also two characteristics relevant to particular counselor/client dyads, the client's score on the predictor, x, and counselor's maximum acceptable error rate for any given client, α. The base rate, ϕ, can be used to find a value of y, say y_{cut}, that cuts off ϕ proportion in the upper tail of the criterion distribution. Using CHT methods, we can find the value of x, x_{crit}, that has $1 - \alpha$ proportion of the area in its conditional distribution beyond y_{cut}. In this setup, we imagine a particular client's x value being compared to x_{crit}, and if $x \geq x_{crit}$ (i.e., the client tests positive), the client is diagnosed as being in the clinical

group, and if $x < x_{crit}$ (i.e., the client tests negative), the client is diagnosed as not being in the clinical group.

An equivalent decision to the one just described would result if the counselor found the probability that a given client was not in the clinical group and compared p_x to α, assigning the client to the clinical group if $p_x < \alpha$.

To relate the above to language used in test interpretation, we will switch to $1 - p_x$ and $1 - \alpha$, classifying the client into the clinical group if $(1 - p_x) \geq (1 - \alpha)$. The reason for this change is that test interpretations are more likely to talk about what a client <u>is</u> (the likelihood the client is in the clinical group) than what a client <u>is not</u> (the likelihood the client is not in the clinical group). Also, $1 - p_x$, in the above model, increases with x, and this will be useful later.

For a client with a score of x and a counselor whose error rate is α, we can think of two 2-by-2 tables, where testing positive or negative is crossed with actual group membership:

Given x, where $x \geq x_{crit}$

	Clinical Group	Not Clinical Group	
Positive	$1 - p_x$	p_x	1
Negative	0	0	0
	$1 - p_x$	p_x	

Given x, where $x < x_{crit}$

	Clinical Group	Not Clinical Group	
Positive	0	0	0
Negative	$1 - p_x$	p_x	1
	$1 - p_x$	p_x	

The zeroes in the above tables represent the fact that above x_{crit}, one cannot test negative, and below x_{crit}, one cannot test positive.

Now let's consider the decision process where instead of being given $1 - p_x$, the counselor is given a non-numeric statement that is intended to mean $1 - p_x$, such as "the client is likely to be depressed." (This assumes that the test interpreter has y_{cut} and ρ for a population relevant to the client. This may be a good-hearted assumption.) At this point, ambiguity enters the picture. The statement was intended to mean $1 - p_x$, but how was it understood? We know from our research (Lichtenberg & Hummel, 1999) and other's (see above references) that there is variability in meaning. Let's assume that the meaning, i.e., the counselor's subjective numerical estimate of the likelihood represented by the statement, is denoted as t and represents a random variable on the interval 0 to 1. Further, let's assume that the mean of this variable is equal to $1 - p_x$ (perhaps another good-hearted assumption). What we want to know is $Pr[\, t \geq 1 - \alpha \mid 1 - p_x]$. What this says is that for clients with a score of x seeing counselors who use an error rate of α, what is the probability that the client will be classified into the clinical group on the basis of the non-numeric statement representing $1 - p_x$? To answer this question, we need to know the distribution of t.

For our model we use the standard form of the beta distribution for t. This distribution is defined on the interval 0 to 1. It is used to approximate the binomial distribution, and the binomial came to mind as we looked at histograms of the data in our earlier study (Lichtenberg & Hummel, 1998). Also, it has been used to model probabilities when they are considered a random variable as they are here.

Now, for a given value of $1 - p_x$ and $1 - \alpha$, we can use the beta distribution to define p_{p_x}, the probability of testing negative for a given value of $1 - p_x$ (or, alternatively, x).

To compute the probability of testing positive and being in the clinical group is straightforward, because these two events are independent for a given value of x. Whether or not one is in the clinical group, given x, depends on $y - \rho x$, a client's error of prediction. Whether or not one tests positive, given x, depends on $t - (1 - p_x)$, a counselor's deviation from the mean probability estimate. The values $y - \rho x$ and $t - (1 - p_x)$ are independent realizations of random variables defined on two distinct populations.

For a client with a score of x, we can think of the following 2-by-2 table:

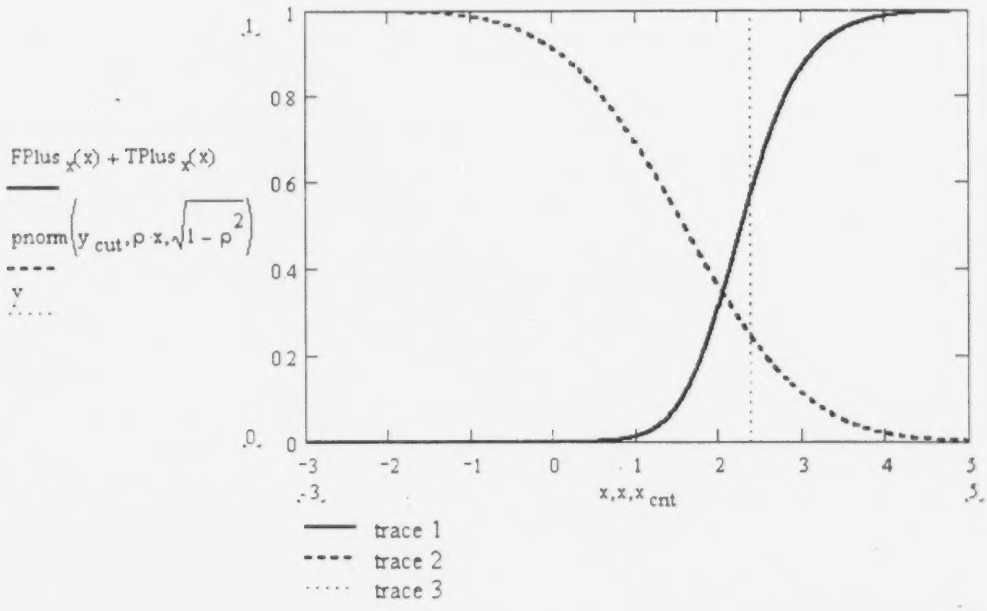
Given x

	Clinical Group	Not Clinical Group	
Positive	$(1-p_{p_x})(1-p_x)$	$(1-p_{p_x})p_x$	$1-p_{p_x}$
Negative	$p_{p_x}(1-p_x)$	$p_{p_x}p_x$	p_{p_x}
	$(1-p_x)$	p_x	

What we can conclude from this table is that whether one tests positive or not is no longer a simple function of x_{crit}. Some clients test positive below this value and some test negative above it.

Because of the independence of the events represented in the table, for a given value of x, the probability that one who tests positive is a false positive is simply $\frac{(1-p_{p_x})p_x}{1-p_{p_x}} = p_x$, the probability that one is not in the clinical group given x. Therefore, for any given value of x, the probability is p_x that the client will be incorrectly classified into the clinical group because the counselor interprets the non-numeric statement as $t \geq 1-\alpha$. While p_x is unaffected by the various interpretations of the non-numeric statement of probability, the probability of testing positive is.

To get an idea of the effect of the ambiguity, we present an example that assumes that the validity coefficient is $\rho = 0.65$, the base rate in the clinical group is $\phi = 0.15$, a counselor's maximum probability of a false positive for any given client is $\alpha = 0.25$, and $\sigma_t^2 = 0.15(\mu_t - \mu_t^2)$ is the variance for the beta distribution with mean $\mu_x = 1 - p_x$.



In the above figure, the dashed line (trace 2) represents p_x, the probability of *not* being in the clinical group, and it decreases as x increases. The solid line (trace 1) represents the probability of testing positive, $1 - p_{p_x}$, and it increases with x. The light dotted vertical line (trace 3) gives the position of x_{cnt}, the point that would have defined positive and negative tests had a numerical probability been given. Without the ambiguity, the probability of testing positive, p_x, would be given by a step function, which would equal zero for values of x less than x_{cnt}, and one otherwise. With ambiguity, it is clear that there would be many false positives that

would not have occurred had numeric information been given. This is especially apparent just to the left of the critical value. For example, if a client had an x of 1.5, then he has a probability of $1 - p_{p_x} = 0.082$ of testing positive.[3] Since the critical value of x for $\alpha = 0.25$ is $x_{crit} = 2.383$, this client could not have tested positive without the ambiguity inherent in the non-numeric presentation of probabilistic information. Having tested positive, the client has a probability of not being in the clinical group of $p_x = 0.532$. We assume this latter probability would disturb many counselors, for they would be better off flipping a coin.

A different perspective might be taken with respect to the current example. If we use the above tabular representation of the probabilities for a client with an x of 1.5, we have the following:

Given x =1.5

	Clinical Group	Not Clinical Group	
Positive	0.038	0.044	0.082
Negative	0.429	0.489	0.918
	0.468	0.532	

(column totals duffer from column sum due to rounding)

The probability of a false positive is 0.044, and a counselor might take the position that since for clients with an x of 1.5, less than 5% would be treated incorrectly, there would be no cause for concern.

The two perspectives are in sharp contrast. One counselor says, "For clients 1.5 standard deviations above the mean, I incorrectly treat only 4.4% of them. No problem." Another counselor says, "For clients I treat who are 1.5 standard deviations above the mean, 53.2% are treated incorrectly. This is disastrous." For reasons spelled out in CHT, and based on the admonition to "first, do no harm," we take the second point of view.

To this point, the discussion has been with respect to a given value of x. We now turn to the implications of the model when x is allowed to vary from $-\infty$ to $+\infty$. To evaluate the model with respect to the range of x, we need to sum each cell probability across the tables corresponding to the values of x, weighting each by the likelihood of x.

Using this line of reasoning, while changing the range of integration and using either p_x or $1 - p_x$, as appropriate, the other probabilities in the table can be computed. The results for our example are as follows:

For x from $-\infty$ to $+\infty$

	Clinical Group	Not Clinical Group	
Positive	0.007069	0.001515	0.008584
Negative	0.142931	0.848485	0.991417
	0.150000	0.850001	

Even given what we consider to be generous values for ρ, ϕ, and α, relatively fewer clients would test positive. Of those that did, 0.001515 / 0.008584 = 0.17653 proportion of them would be incorrectly classified.

Likewise, with ambiguity in the model, to evaluate the model with respect to the range of x, we sum each cell probability across the tables corresponding to the values of x, weighting each by the likelihood of x. Then using the same line of reasoning [and using either $(1-p_{p_x})(1-p_x)$, $(1-p_{p_x})p_x$, $p_{p_x}(1-p_x)$, or $p_{p_x}p_x$, as appropriate], the four long-run probabilities in the table can be computed. The results for our example are as follows:

For x from $-\infty$ to $+\infty$
(including ambiguity)

	Clinical Group	Not Clinical Group	
Positive	0.014851	0.008657	0.023508
Negative	0.135149	0.841312	0.976460
	0.150000	0.849968	

Of those that did test positive, 0.008657 / 0.023508 = 0.368252 proportion of them would be incorrectly classified. The effect of ambiguity is to more than double the probability that those who test positive will be incorrectly classified.

Conclusion

Given the manner in which ambiguity introduces a substantial, independent source of error into the diagnostic process, we suggest that the practice of using non-numeric statements of likelihood be stopped. Instead, results of tests should be presented in tabular form with one row for each diagnostic category. Each row would contain the description of the category, its base rate, the client's probability of membership in the category, and citations to empirical support for the information presented.

Those who provide narrative reports of testing may state that an expert test interpreter, an individual with long experience with the test, can make judgements that have substantial validity. Current trends toward requiring empirical support suggest that such statements not be taken at face value. We believe there is a significant professional need to move toward empirically supported expertise. To do less seems unwise. Why ask counselors to empirically justify their treatments, while not requiring test interpreters to justify their claims in a similar fashion?

References

Bass, B. M., Cascio, W. F., & O'Connor, E. J. (1974). Magnitude estimations of expression of frequency and amount. Journal of Applied Psychology, 59, 313-320.

Beyth-Marom, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. Journal of Forecasting, 1, 257-269.

Brun, W, & Teigen, K. H. (1988). Verbal probabilities: Ambiguous, context-dependent, or both? Organizational Behavior and Human Decision Processes, 41, 390-404.

Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. Organizational Behavior and Human Decision Processes, 36, 391-405.

Butcher, J. N. (1990) MMPI-2 in psychological treatment. New York: Oxford.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. Journal of Consulting and Clinical Psychology, 66, 7-18.

Clarke, V. A., Ruffin, C. L., Hill, D. J., & Beamen, A. L. (1992). Ratings of orally presented verbal expressions of probability by a heterogeneous sample. Journal of Applied Social Psychology, 22, 638-656.

Elliott, R. (1998). Editor's introduction: A guide to the empirically supported treatments controversy. Psychotherapy Research, 8, 115-125.

Foley, B. J. (1959). The expression of certainty. American Journal of Psychology, 72, 614-615.

Goldman, L. (1973). The use of tests in counseling. New York: Appleton-Century-Crofts.

Graham (1993). MMPI-2: Assessing personality and psychopathology (2nd ed.). New York: Oxford.

Green, R. L. (1991). MMPI-2/MMPI: An interpretive manual. Boston: Allyn & Bacon.

Hummel, T. J. (1990). Computer assisted instruction for teaching measurement-based clinical hypothesis testing. Minneapolis: University of Minnesota. (ERIC Document Reproduction Service No. ED 319-756). A paper presented at the annual meeting of the American Educational Research Association, Boston.

Hummel, T. J. (1999). The usefulness of tests in clinical decisions. In J. Lichtenberg & R. Goodyear (Eds.), Scientist-practitioner perspectives on test interpretation (pp. 59-112). Boston, MA: Allyn & Bacon.

Johnson, E. M. (1973). Encoding of qualitative expressions of uncertainty. (Technical Paper No. 250). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Kendall, P. C. (1998). Empirically supported psychological therapies. Journal of Consulting and Clinical Psychology, 66, 3-6.

Lichtenberg, J. W., & Hummel, T. J. (August, 1998). The communication of probabilistic information through test interpretations. Paper presented as a part of a symposium on "Test interpretation in counseling--Recent research and practice" (C. Claiborn, Chair) at the annual convention of the American Psychological Association, August, 1998, San Francisco, CA.

Lichtenstein, S., & Newman, J. R. (1967). Empirical scaling of common verbal phrases associated with numerical probabilities. Psychonomic Science, 9, 563-564.

Ness, M. E. (1995). Ordinal positions and scale values of probability terms as estimated by three methods. Measurement and Evaluation in Counseling and Development, 28, 152-161.

Reagan, R. T., Mosteller, F., & Youtz, C. (1989). Quantitative meanings of verbal probability expressions. Journal of Applied Psychology, 74, 433-442.

Simpson, R. H. (1944). The specific meanings of certain terms indicating differing degree of frequency. The Quarterly Journal of Speech, 30, 328-330.

Simpson, R. H. (1963). Stability in meanings for quantitative terms: A comparison over 20 years. Quarterly Journal of Speech, 49, 146-151.

Sutherland, H. J., Lockwood, G. A., Tritchler, D. L. , Sem, F., Brooks, L., & Till, J. E. (1991). Communicating probabilistic information to cancer patients: Is there "noise" on the line? Social Science and Medicine, 32, 725-731.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Journal of Experimental Psychology: General, 115, 348-365.

Footnotes

[1] At the same time, research evidence suggests (a) that the probabilistic statements shared with clients by counselors are not interpreted or understood uniformly by clients and (b) that the meaning the counselor intends to convey is unlikely to be meaning understood by the client.

[2] Although the discussion here is cast in terms of a single predictor, that predictor could be any linear or non-linear function of a set of independent variables that place an individual on a continuum. Even in the context of configural analysis using the MMPI (e.g., Butcher, 1990; Graham,1993; Green, 1991), one could imaging ordering individuals with respect to their "4-9-ness."

[3] All probabilities were computed using the numerical integration facilities provided by Mathcad 8.0. (MathSoft, Inc. 201 Boradway, Cambridge, MA 02139).